SOX7: Novel Autistic Gene Identified by Analysis of Multi-Omics Data

Samantha Gonzales[1₺], Jane Zizhen Zhao[2₺], Na Young Choi[3], Prabha Acharya[3], Sehoon Jeong[4]
Moo-Yeal Lee[3]*

[1] Department of Biostatistics, Florida International University, Miami, FL 33199
[2] Miami Dade College Kendall Campus and School for Advanced Studies, Miami, FL 33176
[3] Department of Biomedical Engineering, University of North Texas, Denton, TX 76207
[4]Department of Healthcare Information Technology Inje University, Gimhae, South Korea, 50834

[₺]Equivalent contributions

*Corresponding to:
Moo-Yeal Lee, PhD
Associate Professor
Department of Biomedical Engineering, University of North Texas, Denton, TX 76207
Email: Moo-Yeal.Lee@unt.edu

ABSTRACT

Background: Genome-wide association studies and next generation sequencing data analyses based on DNA information have identified thousands of mutations associated with autism spectrum disorder (ASD). However, more than 99% of identified mutations are non-coding. Thus, it is unclear which of these mutations might be functional and thus potentially causal variants. Transcriptomic profiling using total RNA-sequencing has been one of the most utilized approaches to link protein levels to genetic information at the molecular level. The transcriptome captures molecular genomic complexity that the DNA sequence solely does not. Some mutations alter a gene's DNA sequence but do not necessarily change expression and/or protein function. To date, few common variants reliably associated with the diagnosis status of ASD despite consistently high estimates of heritability. In addition, reliable biomarkers used to diagnose ASD or molecular mechanisms to define the severity of ASD do not exist.

Objectives: It is necessary to integrate DNA and RNA testing together to identify true causal genes and propose useful biomarkers for ASD.

Methods: We performed gene-based association studies with adaptive test using genome-wide association studies (GWAS) summary statistics with two large GWAS datasets (ASD 2019 data: 18,382 ASD cases and 27,969 controls [discovery data]; ASD 2017 data: 6,197 ASD cases and 7,377 controls [replication data]) which were obtained from the Psychiatric Genomics Consortium (PGC). In addition, we investigated differential expression for genes identified in gene-based GWAS with a RNA-seq dataset (GSE30573: 3 cases and 3 controls) using the DESeq2 package.

Results: We identified 5 genes significantly associated with ASD in ASD 2019 data (KIZ-AS1, $p=8.67\times10^{-10}$; KIZ, $p=1.16\times10^{-9}$; XRN2, $p=7.73\times10^{-9}$; SOX7, $p=2.22\times10^{-7}$; PINX1-DT, $p=2.14\times10^{-6}$). Among these 5 genes, gene SOX7 (p=0.00087), LOC101929229 (p=0.009), and KIZ-AS1 (p=0.059) were replicated in ASD 2017 data. KIZ (p=0.06) was close to the boundary of replication in ASD 2017 data. Genes SOX7 (p=0.0017, adjusted p=0.0085), LOC101929229 (also known as PINX1-DT, $p=5.83\times10^{-7}$, adjusted $p=1.18\times10^{-5}$), and KIZ (p=0.00099, adjusted p=0.0055) indicated significant expression differences between cases and controls in the RNA-seq data. SOX7 encodes a member of the SOX (SRY-related HMG-box) family of transcription factors pivotally contributing to determining of the cell fate and identity in many lineages. The encoded protein may act as a transcriptional regulator after forming a protein complex with other proteins leading to autism.

Conclusion: Gene SOX7 in the transcription factor family could be associated with ASD. This finding may provide new diagnostic and therapeutic strategies for ASD.

INTRODUCTION

Autism spectrum disorder (ASD) is a heterogeneous grouping of neurodevelopmental traits which is diagnosed in roughly 1% of the world population (Fombonne 2009). ASD conditions are characterized by having attention-deficit hyperactivity disorder (ADHD), intellectual disability (ID), epilepsy, social communication deficits and restricted, repetitive, or unusual sensory-motor behaviors, or gastrointestinal problems (Gillberg, Fernell, and Minnis 2014). A lot of research efforts have gone into understanding the causes of individual differences in autistic behavior. Twin and family studies strongly demonstrate that autism has a particularly large genetic basis, with estimated heritability ranging from 40% to 90% (Gaugler et al. 2014; Devlin et al. 2013; Tick et al. 2016; Sandin et al. 2017). Molecular genetic studies revealed that the genetic risk for autism is shaped by a combination of rare and common genetic variants (Consortium 2017).

Over the past decade, genome-wide association (GWAS) and other type of genetic studies have identified increasing numbers of single nucleotide polymorphisms (SNPs) (Grove et al. 2019; Robinson et al. 2016) and other forms of genetic variation that are associated with ASD (Bourgeron 2015). It has been estimated that more than 100 genes and genomic regions are associated with autism (Sanders et al. 2015; Satterstrom et al. 2020). While most of these studies focused on identifying heritable single nucleotide polymorphisms (SNPs) associated with ASD risk, other studies have demonstrated the influence of de novo mutations ranging from a single base (Sanders et al. 2012; O'Roak et al. 2012) thousands to millions of bases long (Sanders et al. 2011; Levy et al. 2011) such as to copy number variants (CNVs). Several likely gene-disruptive (LGD) variants in genes such as GRIK2 (Jamain et al. 2002) and ASMT (Melke et al. 2008) affecting autism-risk were found exclusively or more frequently in individuals with autism compared to control groups. Jamain et al.(Jamain et al. 2003) showed strong evidence suggesting that mutations in NLGN3 and NLGN4 are involved in autism spectrum disorder. Additionally, deletions at Xp22.3 that include NLGN4 have been reported in several autistic individuals. Roohi et al. found out that (Roohi et al. 2009) CNTN4 plays an essential role in the formation, maintenance, and plasticity of neuronal networks. Disruption of CNTN4 is known to cause developmental delay and mental retardation. This report suggests that mutations affecting CNTN4 function may be relevant to ASD pathogenesis. A review by Li and Brown (Li, Zou, and Brown 2012) discussed a substantial body of evidence has resulted from genome-wide screening for several widely studied candidate ASD genes. Similarly, a large-scale international collaboration was conducted to combine independent genotyping data to improve statistical power and aid in robust discovery of GWS loci (Consortium 2017). This international collaboration also identified a significant genetic correlation between schizophrenia and autism with several neurodevelopmental related genes such as EXT1, ASTN2, MACROD2, and HDAC4. A combined analysis investigating both rare and common gene variants supported the evidence of the role of several genes/loci associated with autism (e.g., NRXN1, ADNP, 22q11 deletion) and revealed new variants in known autism-risk genes such as ADPNP, NRXN1, NINL, MECP2 and identified new compelling candidate genes such as KALRN, PLA2G4A, and RIMS4 (Leblond et al. 2019). Recently, Buxbaum (Buxbaum 2022) summarized the prevalence of some genetic variants in subjects ascertained for ASD.

Research investigating the gene expression profiles of those with ASD has also proven insightful genetic contributions to ASD. Expression levels of genes containing rare mutations associated with autism were evaluated in lymphoblasts from autism cases and controls, including aforementioned genes such as NLGN3, NLGN4, NRXN1, and MeCP2. Out of these, NLGN3 was found to be differentially expressed along with SHANK3 (Yasuda et al. 2011). More comprehensive gene expression analyses have confirmed susceptibility genes previously reported in GWAS-based analysis, identified novel differentially expressed genes, and biological pathways enriched for these genes (Rahman et al. 2020). RNA sequencing data analyses have elucidated several potential drivers of autism susceptibility, such as resting-state functional brain activity (Berto et al. 2022), dopaminergic influences in the dorsal striatum (Brandenburg et al. 2020), overexpression of FOXP1, a gene involved in regulating tissue and cell type specific gene transcription in the brain (Chien et al. 2013; Ferland et al. 2003), and genome-wide alterations to lncRNA levels, downregulation of alternative splicing events, and brain-region dependent alterations in gene expression (Parikshak et al. 2016). Aforementioned studies indicate that integrating GWAS and RNA-seq data analysis can provide a better picture of the various underlying mechanisms behind a heterogeneous, multifaceted disorder like ASD.

In this study, we performed whole genome gene-based association tests for ASD with the adaptive test method (Guo and Wu 2018) using summary statistics from two large GWAS datasets which were obtained from the Psychiatric Genomics Consortium (PGC). We identified 5 genes significantly associated with ASD in ASD 2019 data. Among these 5 genes, gene SOX7 was replicated in ASD 2017 data. Further RNA sequencing data analysis indicated that gene SOX7 was significantly upregulated in cases as compared to controls. SOX7 encodes a member of the SOX (SRY-related HMG-box) family of transcription factors pivotally contributing to determining of the cell fate and identity in many lineages. The encoded protein may act as a transcriptional regulator after forming a protein complex with other proteins leading to autism.

METHODS

Datasets

Discovery GWAS summary statistics: The discovery dataset (labeled as asd2019) includes summary statistics from a meta-analysis of European samples derived from two cohorts: a population-based case control study from the iPSYCH project and a family trio-based study from the Psychiatric Genomics Consortium (PGC) (Grove et al. 2019). The iPSYCH samples included individuals born to a known mother and a resident of Denmark at the time of their first birthday. Cases were identified using the Danish Psychiatric Central Research Register, using diagnoses from 2013 or earlier by psychiatrist according to diagnostic code ICD10, which includes diagnoses of childhood autism, atypical autism, Asperger's syndrome, "other pervasive developmental disorders", and "pervasive developmental disorder, unspecified" (Grove et al. 2019). The PGC samples consisted of 5 cohorts, whose trios were analyzed as cases and pseudo-controls. Details regarding these studies can be found in (Grove et al. 2019) and (Consortium 2017). The combined sample size consisted of 18,382 cases and 27,969 controls. Imputation and quality control were performed *via* PGC's Ricopili pipeline, which ensures to produce robust, reproducible, and comparable datasets. The iPSYCH samples were processed separately in the 23 genotyping batches, while the PGC samples were processed separately for each study. Genotype imputation was performed with IMPUTE2/SHAPEIT(Bryan N. Howie 2009; Delaneau, Marchini, and Zagury 2011) in the Ricopili pipeline using the 1000 Genomes Project phase 3 dataset as the reference set. Regions demonstrating high linkage disequilibrium were excluded, and one of high similarity pairs of subjects identified by PLINK's identity by state (IBS) analysis (Chang et al. 2015) were reduced at random, with a preference for retaining cases. Association was performed using PLINK on imputed dosage data and the meta-analysis was performed using METAL (Grove et al. 2019). More detailed descriptions of each stage of the analysis can be found in Grove et al (Grove et al. 2019). The summary statistics produced by this study and subsequently used for our analysis can be found at https://pgc.unc.edu/for-researchers/download-results/.

Replication GWAS summary statistics: The replication dataset (labeled as asd2017) includes summary statistics from a European-ancestry meta-analysis performed by the Autism Spectrum Disorders Working Group (AWG) of The Psychiatric Genomics Consortium (PGC), which aimed at improving statistical power to detect loci significantly associated with ASD. The meta-analysis was performed on data from 14 independent cohorts across different ancestries totaling over 16,000 individuals. For each step in the meta-analysis, each cohort was processed individually. Individuals were excluded if they were assessed at less than 36 months of age or if diagnostic criteria were not met from the Autism Diagnostic Interview-Revised (ADI-R) or the Autism Diagnostic Observation Schedule (ADOS) domain scores. While a “world-wide” meta-analysis on this aggregate dataset was performed, we derive our replication dataset based on the smaller European-only analysis consisted of 6,197 ASD cases and 7,377 controls (Consortium 2017). Each stage of the imputation and quality control was performed similarly as the asd2019 data: Imputation and quality control on PGC samples were performed following the PGC’s “Ricopili” pipeline. Since multiple studies were involved, necessary studies were performed to check for and remove duplicate individuals prior to imputation. Family trio-based data was organized as case and pseudo-controls. Criteria for SNP retention and other pre-imputation quality control steps can be found in the study’s supplementary File 1 (Consortium 2017). Genotype imputation was performed with IMPUTE2/SHAPEIT using the 2,184 phased haplotypes from the full 1000 Genomes Project dataset as the reference set. All 14 cohorts were tested for association individually using an additive logistic regression model in PLINK. More detailed information about each stage of the analyses performed by this study can be found in the study’s supplementary File 1 (Consortium 2017). The resulting summary statistics which were utilized in our analysis can be found at https://pgc.unc.edu/for-researchers/download-results/.

Bulk RNA-Seq: The RNA dataset was obtained from a gene co-expression analysis which aimed to identify modules of co-expressed genes associated with ASD (I et al. 2011). The study can be found in the Gene Expression Omnibus (GEO) database, under accession number GSE30573. Detailed descriptions of the raw data acquisition and quality control processes can be found in the supplementary information of (I et al. 2011) as well as the GEO accession viewer. Briefly, brain tissue samples (frontal cortex, temporal cortex, and cerebellum) were obtained from the Autism Tissue Project (ATP) and the Harvard Brain Bank. Cases were diagnosed using ADI-R diagnostic scores, which can be found along with other clinical data upon request from the ATP website. Total RNA was extracted from the sample tissues following the Qiagen miRNA kit instructions. Quality and concentration were assessed by Agilent Bioanalyzer and Nanodrop, respectively. Reads were generated using Illumina GAII sequencer using manufacturer settings and were 73-76 nucleotides in length. Raw sequencing data for the frontal and temporal cortex samples were available in the SRA run selector for 6 autism cases and 6 controls. Out of the 12 total samples, 8 were temporal cortex samples (4 cases and 4 controls) and 4 were frontal cortex samples (2 cases and 2 controls) (I et al. 2011).

Quality Control & Preprocessing

GWAS summary data: After downloading the raw summary statistics from the PGC website, we performed quality control analysis to ensure robust and quality results. Only SNPs on autosomal chromosomes were used. First, SNPs with an imputation information metric (INFO) score > 0.9 were removed. Next, SNPs with strand-ambiguous alleles or non-biallelic loci were removed as well as SNPs with duplicate rsIDs. Z scores were then calculated using each variants odd’s ratio and standard error using the equation $Z = \log(OR)\,/SE(\log\,(OR))$. After quality control, the raw variants were sorted into hg19 RefSeq genes. Linkage disequilibrium (LD) within each gene was calculated using the 1000 Genomes European reference panel (phase 3): For each gene, a subset of the GWAS variants $\pm 1000\ bp$ of the gene's transcription start site and transcription end site were matched to the reference variants, ensuring both used the same reference allele and flipping Z score signs if necessary. Genes that contained less than 2 SNPs were removed. The Pearson's correlation between this subset of genotypes was calculated and used as the gene-wide LD. One SNP of a pair of SNPs with perfect correlation ($r_{ij} = 1$) within a gene was removed. The processed data was saved in 22 'RData' files (one for each chromosome) containing a list of data-frames, where each list element comprised of 1) SNP information for a specific gene and 2) its corresponding LD matrix.

RNA-seq data: The sequence read archive (SRA) accession list and associated sample metadata ("SRA Run Table") for GSE30573 were downloaded from the SRA run selector page for the study. Raw fastq files were downloaded from the SRA using the SRA Toolkit *via* the 'prefetch' and 'fastq-dump' commands ("SRA Toolkit"). We used FastQC to assess the quality of reads in each file, and MultiQC to visualize the results in batch format (Andrews 2010; Ewels et al. 2016). Only 1 sample failed the 'per sequence base quality' assessment and was subsequently trimmed of low-quality reads using the command-line tool 'fastq_quality_filter' from the FastX toolkit using a minimum quality Phred score of 20 and a minimum percent of bases per read to meet that threshold of 50% ("FASTX-Toolkit"). Reassessment *via* FastQC demonstrated this as sufficient trimming to meet the quality needed for downstream analysis.

Statistical Analysis

Gene-based Association Test: To perform gene-based association testing, we used the function 'sats' in the R package 'mkatr' (Guo and Wu 2018). This function computes p-values for 3 different SNP-set testing methods using GWAS summary statistics and an LD matrix calculated from a reference panel. A brief description of each is as follows: Let m denote the number of variants considered in a gene or gene region and let $(z_1, \cdots z_m)$ represent the GWAS summary statistics for each. Let $R = (r_{ij})$ denote the estimated correlation between Z statistics based on variant linkage disequilibrium (LD) calculated from a reference panel (Guo and Wu 2018). The tests included in the sats function are the sum test (a type of burden test), the squared sum test (a type of SKAT statistic) and the adaptive test (similar to the SKAT-O statistic). The three tests are as follows:

1. Sum test (ST): $B = \sum_{j=1}^{m} z_j$
2. Squared sum test (S2T): $Q = \sum_{j=1}^{m} z_j^2$
3. Adaptive test (AT): $T = \min_{\rho \in [0,1]} P\left(Q_\rho\right), where\ Q_\rho = (1-\rho)Q + \rho B^2\ and\ P(Q_\rho)$ denotes its p-value.

It can be shown that Q_ρ asymptotically follows a weighted sum of independent chisqured distribution with 1 degree freedom $[\chi^2\ (df = 1)]$ whose weights equal the eigenvalues of R, allowing for efficient computation of $P\left(Q_\rho\right)$. The minimum p-value of AT is searched for over a range of ρ in the interval [0, 1] (Guo and Wu 2018).

The ST is most valuable when all variants have the same direction of effect and approximately equal effect size, while the S2T will perform better than ST when variants have different effect directions. AT utilizes information from both ST and S2T, meaning AT can adapt to the variants in the data better than ST or S2T alone. Indeed, the adaptive test shows the most robust performance across a wider range of scenarios (Guo and Wu 2018). More details regarding the derivation of these tests and their relation to the single-variant association test can be found in (Guo and Wu 2018).

Differential Expression Analysis: After passing quality control, RNA-seq reads were aligned to a reference genome using STAR (Dobin et al. 2013) by following two steps: genome indexing and the alignment to the indexed reference genome. We generated the genome index files using STAR's –genomeGenerate flag and setting –sjdbOverhang to 75 to match the maximum read length - 1 across the samples. The reference genome FASTA file and corresponding annotation GTF file (GChr37/hg19, release 41) used to generate these index files were downloaded from GENCODE. After alignment, we used HTSeq (Putri et al. 2022) to estimate the number of reads per gene region.

For genes with gene expression counts at least 10, we used the R package DESeq2 (Love, Huber, and Anders 2014) to perform differential expression analysis based on normalized gene expression counts. DESeq2 uses a generalized linear model to model the relationship between a trait and the log2 fold changes in gene expression (Love, Huber, and Anders 2014). We used the adjusted p-value to assess significance in gene differential analysis.

Computing Environment and Script Availability: RNA-seq quality control, alignment, and counts were processed on the lonestar6 high-performance cluster provided by TACC at the University of Texas at Austin. Differential expression analysis and gene-based association tests were performed in a local Linux (Windows Linux Subsystem) environment using R in RStudio.

RESULTS

Gene-based Association Test

Discovery GWAS: Out of approximately 19,000 genes tested for association with ASD, 5 genes were identified as significant with Bonferroni corrected p-values less than $p = 2.5 \times 10^{-6}$ (**Figure 1** and **Table 1**). SOX7 ($p = 2.22 \times 10^{-7}$) encodes a transcription factor involved in regulating embryonic development and cell fate determination (Takash et al. 2001). KIZ ($p = 1.16 \times^{-9}$) encodes "Kizuna centrosomal protein", which plays a central role in stabilizing the pericentriolar region before the spindle formation step in cellular division (Oshimori, Ohsugi, and Yamamoto 2006). A gene region which encodes long non-coding antisense RNA for KIZ, KIZ-AS1 ($p = 8.67 \times 10^{-10}$) was also identified as significant, however the function of this antisense RNA has not been determined. XRN2 ($p = 7.73 \times 10^{-9}$) encodes a 5'-3' exoribonuclease which is pertinent in promoting transcriptional termination (Eaton and West 2018). Finally, LOC101929229 ($p = 2.14 \times 10^{-6}$), also known as PINX1-DT, is a lncRNA that is considered a "divergent transcript" of the protein coding gene PINX1. While the divergent transcript function is not defined, PINX1 encodes a protein that enables telomerase RNA binding and inhibitor activity and is involved in several related processes, including DNA biosynthesis and protein localization (Johnson 2011).

Replication GWAS: Among these 5 genes, gene SOX7 (p=0.00087), LOC101929229 (p=0.009), and KIZ-AS1 (p=0.059) were replicated in ASD 2017 data. KIZ (p=0.06) was close to the boundary of replication in ASD 2017 data (**Table 1**).

Table 1. Genetic features of significant genes identified in the discovery dataset (significance level=2.5x10^{-6}).

Gene	Chr.	Discovery (asd2019)	Replication (asd2017)	RNA-seq (GSE30573)

		p-value	**p-value**	**LFC**	**p-value**	**Adj. p-value**
SOX7	8	2.22E-07	0.0009	1.17	0.0017	0.0085
LOC101929229	8	2.14E-06	0.009	3.22	$5.83\text{x}10^{-7}$	$1.18\text{x}10^{-5}$
XRN2	20	7.73E-09	0.10	0.34	0.001	0.007
KIZ	20	1.16E-09	0.06	0.63	0.001	0.006
KIZ-AS1	20	8.67E-10	0.059	0.062	0.43	0.58

Abbreviation: Chr.: chromosome; LFC: log2 fold change.
LOC101929229 is also known as PINX1-DT.

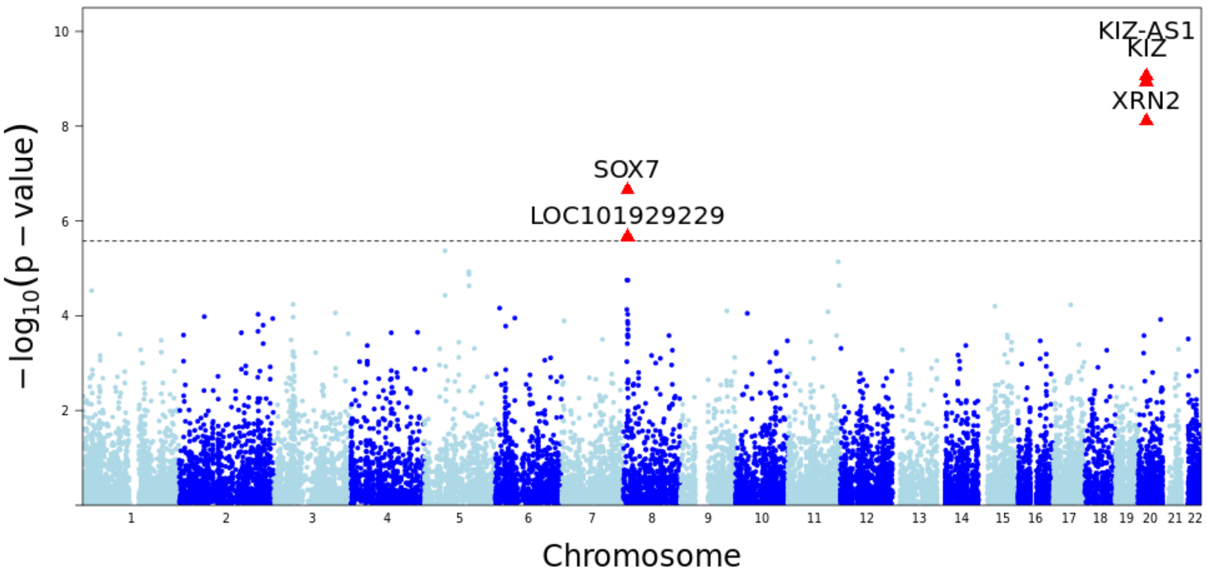


Figure 1. Manhattan plot of the asd2019 GWAS data. Each dot represents a gene tested for association with ASD; the dotted horizontal line represents a Bonferroni corrected p-value threshold of $2.5\text{x}10^{-6}$.

Differential Expression Analysis

Among the five gene identified in the discovery of GWAS, gene SOX7 (log2FoldChange [LFC] = 1.17, $p = 0.0017$; Benjamini-Hochberg (BH) adjusted p = 0.0085), LOC101929229 (LFC=3.22, p=$5.83\times^{-7}$, adjusted p=$1.18\times^{-5}$), KIZ (LFC=0.63, p=0.00099, BH adjusted p=0.0055) were also identified as significant in the differential gene expression analysis (**Table 1**). A comparison of case-control gene expression counts for SOX7 can be found in **Figure 2**, demonstrating that SOX7 is consistently upregulated in autism cases compared to controls. The expression of SOX7 is increased in autism patients relative to controls by a multiplicative factor of 2.25. In addition, the expression of LOC101929229 is increased in autism patients than in controls by a multiplicative factor of 9.31.

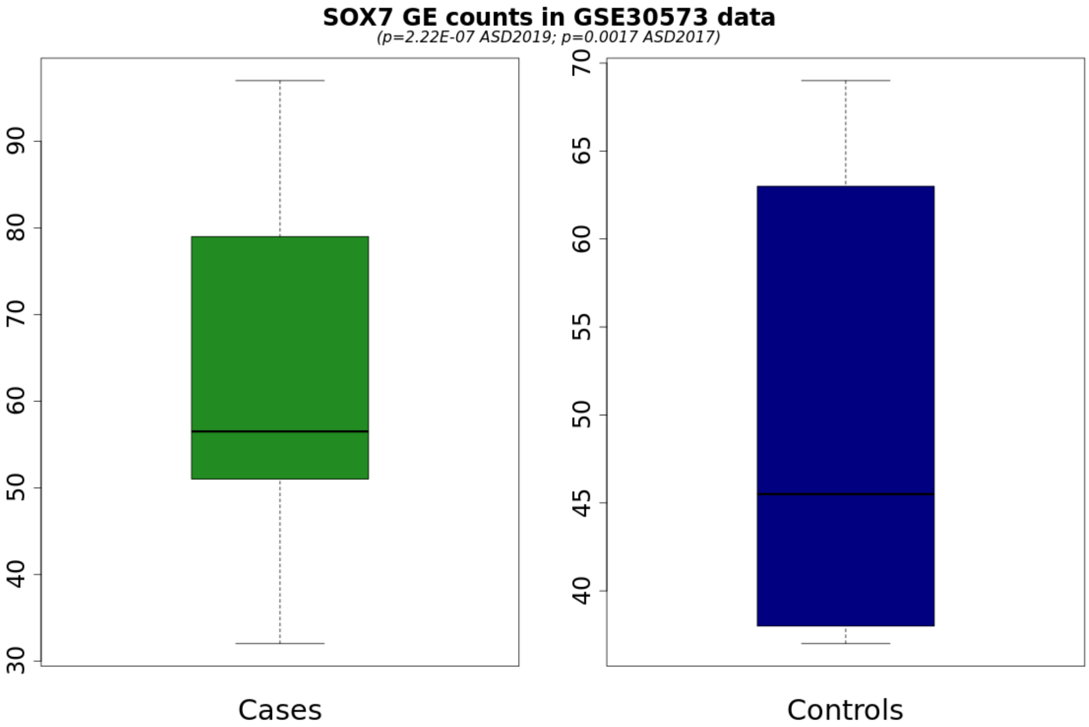


Figure 2. Boxplot demonstrating significant upregulation of SOX7 in autism cases (p = 0.0017).

DISCUSSION

Through gene-based analysis, we identified 5 gene regions (KIZ, KIZ-AS1, XRN2, LOC101929229, and SOX7) significantly associated with ASD. KIZ, LOC101929229, and SOX7 were supported by results from the replication study in a different GWAS data and the differential expression analysis performed on publicly available RNA-seq data.

KIZ is located on chromosome 20, and encodes Kizuna centrosomal protein, which aids in stabilizing the pericentriolar region of centrosomes before spindle formation. KIZ has been identified as significantly associated with autism in previous GWAS (Grove et al. 2019), TWAS (Huang et al. 2021), gene based analysis (Alonso-Gonzalez et al. 2019), and methylation-based studies (Hannon et al. 2018), and the involvement of cell cycle regulation in autism susceptibility has also been implicated in previous research (Pramparo et al. 2015; Packer 2016). KIZ has also been found to be a potentially shared genetic loci between ASD and attention-deficit hyperactivity disorder (ADHD), providing support for its involvement in neurological disorders (Baranova et al. 2022).

XRN2 is located next to KIZ and encodes a 5’-3’ exonuclease that is involved in myriad RNA management processes, including transcriptional termination, miRNA expression regulation, nonsense-mediated mRNA decay, and rRNA maturation (West, Gromak, and Proudfoot 2004; Wang and Pestov 2011; Nagarajan et al. 2013; Brannan et al. 2012). XRN2 has been found to play a role in regulating miRNA expression in neurons specifically, and altered miRNA expression regulation has been investigated as a potential mechanism for autism susceptibility (Kinjo et al. 2013; Hicks and Middleton 2016; Ghahramani Seno et al. 2011; Wu et al. 2016; Abu-Elneel et al. 2008). Likewise, disruption of proper RNA metabolism as a result of altered expression of RNA binding proteins has been implicated in neurological disease as a whole, and the XRN gene family is involved in nonsense-mediated decay of mRNA, a process that has been implicated in autism pathophysiology (Nussbacher et al. 2019; Marques et al. 2022). Previous GWAS have reported SNPs in the region containing XRN2 to be significantly associated with ASD, affirmed by gene-based analysis using MAGMA (Grove et al. 2019). Additionally, a transcriptome-wide association study (TWAS) found XRN2 to be significantly upregulated in autism, in accord with our findings (Pain et al. 2019). Another gene-based analysis found XRN2 to be associated with ASD and upon further investigation *via* gene-network analysis and enrichment analysis found that not only does XRN2 interact with several genes in the cAMP signaling pathway and RNA transport network, but that the enriched KEGG/GO terms for XRN2 (spliceosome, RNA transport, and nucleic acid binding) found to be associated with ASD are also essential processes pivotal to early development (Alonso-Gonzalez et al. 2019). The extensive involvement of XRN2 in such complex mechanisms of gene expression regulation, particularly in neuronal cell types, offer possible insights into the vast heterogeneity of ASD and its overlap with other neurodevelopmental disorders. In fact, more recent research efforts have focused on ascertaining genetic commonalities between ASD and related disorders such as ADHD, obsessive compulsive disorder (OCD), and Tourette Syndrome, of which XRN2 seems to be a shared significant locus (Peyre et al. 2020; Yang et al. 2021).

SOX7 is of particular interest due to its hallmark involvement in the regulation of Wnt/β-catenin pathway **(Figure 3)**, an important developmental signaling pathway. SOX7 and its related SOX family genes encode transcription factors that are critical to the downregulation of the canonical Wnt/β-catenin signaling pathway, which controls embryonic development, adult homeostasis, and is involved in a multitude of cellular processes (Katoh 2002; MacDonald, Tamai, and He 2009). While the Wnt pathway is ubiquitous to nearly all tissue types, proteins involved in Wnt signaling in the brain specifically have been found to localize in the synapses and influence synaptic growth, and knockout murine models of ASD risk genes that are a part of the Wnt pathway have provided support for the disruption of this pathway in autism-like behaviors (Kwan, Unda, and Singh 2016). Indeed, the Wnt/β-catenin signaling pathway has been suggested as a possible avenue for autism pathogenesis in several studies (de la Torre-Ubieta et al. 2016; Kwan, Unda, and Singh 2016; Quesnel-Vallières et al. 2019; Hormozdiari et al. 2015; Vallée, Vallée, and Lecarpentier 2019; El Khouri et al. 2021; Caracci et al. 2021).

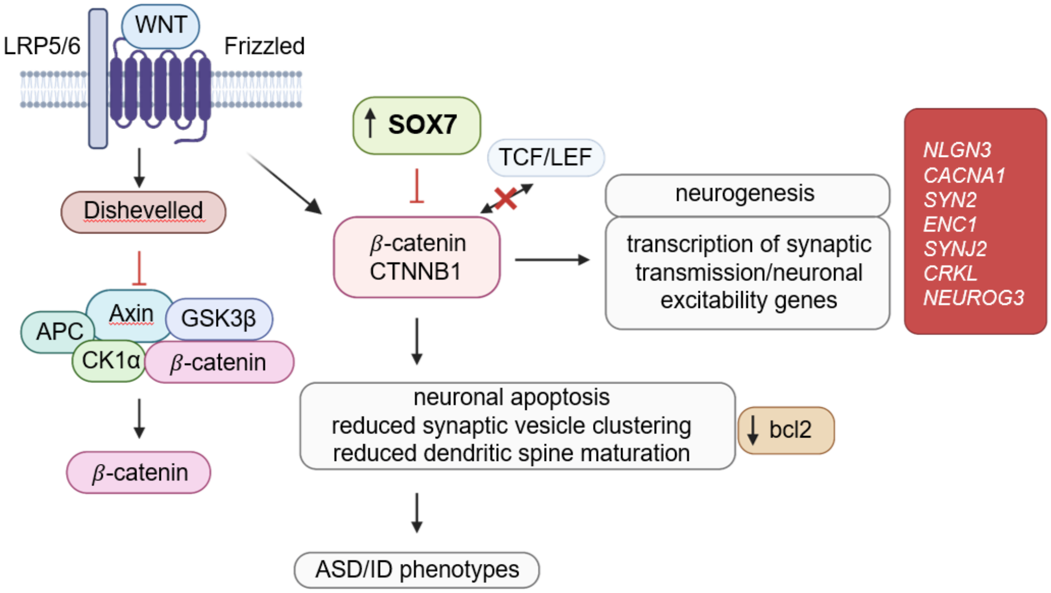

LRP5/6
WNT
Frizzled
↑ SOX7
TCF/LEF
Dishevelled
β-catenin
CTNNB1
neurogenesis
transcription of synaptic transmission/neuronal excitability genes
NLGN3
CACNA1
SYN2
ENC1
SYNJ2
CRKL
NEUROG3
APC
Axin
GSK3β
CK1α
β-catenin
β-catenin
neuronal apoptosis
reduced synaptic vesicle clustering
reduced dendritic spine maturation
↓ bcl2
ASD/ID phenotypes

Figure 3. Proposed pathogenesis of Autism via SOX7.

SOX7 also regulates angiogenesis, vasculogenesis, and endothelial cell development, and the SOX family of transcription factors are critical to cardiovascular development (Francois, Koopman, and Beltrame 2010; K. Kim et al. 2016). For example, SOX7 was found to be upregulated in sustained hypoxic environments, mediating angiogenesis (Klomp et al. 2020), and a knockout model of SOX7 was found to result in profound vascular defects, and demonstrated that SOX7 has an essential role in vasculogenesis and angiogenesis in early development (Lilly et al. 2017). Links between SOX7's role in developmental delay and congenital heart disease have been investigated. Specifically, deletions in the region where SOX7 resides have been demonstrated to simultaneously cause congenital heart defects and intellectual disability (Wat et al. 2009; Páez et al. 2008).

Additionally, Wnt signaling has been demonstrated to orchestrate differentiation of neural vasculature, such as the blood-brain barrier (Stenman et al. 2008; Reis and Liebner 2013). Likewise, there is evidence of vascular involvement in the development of autism (Yao et al. 2006; Ouellette et al. 2020; Emanuele et al. 2010; Casanova 2007). One review in particular suggests that mutations affecting the delicate interactions between Wnt signaling and Shh pathways may alter blood brain barrier integrity in autism by aberrantly interacting with neurovascular molecules (Gozal et al. 2021).

Lastly, oxidative stress has been researched as a potential source of autism susceptibility (Chauhan and Chauhan 2006; Bjørklund et al. 2020), and the interaction between altered vasculature and autism during oxidative stress could point to another potential source of pathogenesis (Yao et al. 2006). Indeed, the role of Wnt/β-catenin signaling in oxidative stress has been implicated in autism susceptibility directly (Zhang et al. 2012). This combination of evidence that implicates both Wnt signaling and SOX7 interactions in the multitude of interrelated processes that have been suggested as mechanisms behind the etiology of ASD, supplemented by our findings, provide ever-mounting support for more in-depth investigations of these particular genes and pathways.

Wnt/β-catenin, oxidative stress, and impaired/altered vasculature have all been implicated in the development of ASD. These three factors are involved with each other and multiple systemic processes, which may contribute to ASD's symptom heterogeneity. The fact that SOX7 is involved in the regulation of Wnt/β-catenin and vasculogenesis points to a potential converging mechanism behind the pathophysiology of ASD. Additionally, the association of SOX7 with autism has been investigated directly. A case study involving a child patient exhibiting "8p23.1 duplication syndrome", revealed a de novo 1.81 Mbp duplication event on chromosome 8 (8p23.1), spanning the region where SOX7 lies (Weber et al. 2014). This patient exhibited characteristic symptoms of the disorder, including delay of motor and speech development and intellectual disability, which heavily overlap with autism and related intellectual disorders. Indeed, this patient also exhibited symptoms specific to ASD, such as repetitive compulsive behavior.

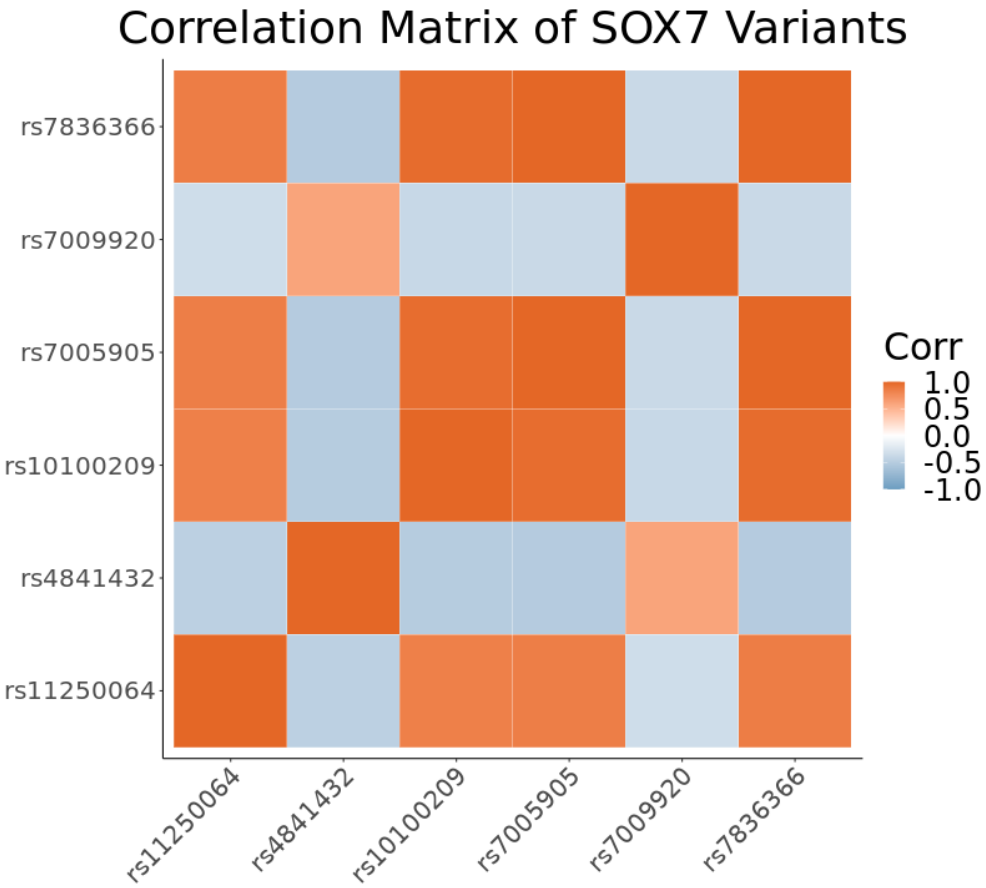
A GWAS performed in a Mexican population found that SOX7 was differentially methylated between autism cases and controls (Aspra et al. 2022). Another study also found that differential methylation was associated with an "elevated polygenic burden" for autism, and further identified that two significantly associated CpG sites were located near GWAS markers for autism on chromosome 8, in the same region as SOX7 (Hannon et al. 2018). It is worth noting that this study also found evidence of SNPs associated with both autism and DNA methylation that were annotated to KIZ and XRN2, two genes that we also found to be significantly associated with ASD.

Changes in methylation lead to changes in gene expression, providing another plausible mechanism of SOX7 involvement: a change in SOX7 methylation affects the expression and thus availability of the transcription factor it encodes, which has a downstream effect on the subsequent pathways SOX7 regulates, such as Wnt/β-catenin. Indeed, both methylation studies demonstrated a negative difference in methylation between autism cases and controls. Generally speaking, undermethylation results in a less compact 3-dimensional genome structure, allowing for greater access to the gene and an increase in expression, which we see evidence of in the higher gene expression counts in autism cases versus controls in our RNA-seq data (**Figure 2**) (Lewis and Bird 1991; Keshet, Lieman-Hurwitz, and Cedar 1986; Buitrago et al. 2021).

Finally, altered expression of SOX7 has been shown to play a role in the development of different types of gliomas. One study demonstrated SOX7 to be downregulated in human glioma, allowing cancer development through upregulated Wnt/β-catenin signaling (Zhao et al. 2016), whereas another study demonstrated that overexpression of SOX7 in high-grade glioma (HGG) promoted cancer development by promoting tumor growth *via* vessel abnormalization (I.-K. Kim et al. 2018). These somewhat conflicting observations demonstrate that, due to its heavy involvement in regulating several intricately linked developmental and homeostatic functions, SOX7 expression must be delicately balanced. Interestingly, it has also been demonstrated that there is extensive overlap of genetic risk between autism and cancer (Crawley, Heyer, and LaSalle 2016; Gabrielli, Manzardo, and Butler 2019; Crespi 2011; Tabarés-Seisdedos and Rubenstein 2009). SOX7 expression and its interactions may provide additional support to this conjecture, particularly due to its role in vasculature development and Wnt signaling regulation.

The methods performed are not without limitations. Gene expression is a very dynamic process that is not only tissue dependent, but also cell type specific and varies depending developmental stage and even external factors (Shen-Orr et al. 2010; Lawlor et al. 2017; Weyer and Schilling 2003; Xu et al. 2014; Fitzgerald et al. 2004; Hsieh et al. 2000). Certainly, these factors affecting genetic expression means that any autism-related genes which are differentially expressed at different development stages or other varying contexts may be missed. Additionally, differential expression analysis was performed on bulk-RNA, whereas it is possible that altered gene expression between autism cases and controls is cell-type specific; knowing the specifics of the expression state of specific cell types that make up key areas of the brain have a better chance of revealing mechanisms behind autism pathogenesis as well as possibly elucidate the pathophysiology behind the vast variety of ASD subtypes. Gene-based analysis also has some limitations, the most important being the reliance on a reference population for estimating linkage disequilibrium between variants. The similarity of this reference population to the population of study is crucial to the accuracy of many gene-based analyses including those performed here. As a result, the extent of our findings is limited to European populations, as this was our reference of choice. Future steps include a tighter integration of DNA and RNA information as well as extensions to non-European populations that have been under-researched.

These limitations notwithstanding, the study has considerable strengths. The AT method used in the gene based GWAS study can not only integrate the good properties of sum and squared sum tests but also consider LD information among genetic variants. The heatmap of the correlation between genetic variants in SOX7 (**Supplementary Figure 1**) indicates that rs7005905 and rs7836366, rs10100209 and rs7836366, and rs10100209 and rs7005905 have strong positive linkage disequilibrium (LD) (ρ>0.5); rs4841432 has negative LD with other variants except for rs7009920. The strong LD in SOX7 and the powerful AT method warrant our identification of the autism associated gene SOX7. The successful replications of SOX7 in the replication data, gene expression data, and the associated biological plausibility underscores the robustness of the finding of the connection between SOX7 and autism. This finding may significantly advance our understanding of the etiology of autism, open new opportunities to reinvigorate the stalling autism drug development and increase the accuracy of risk prediction of autism which makes autism early intervention and prevention being possible.



Supplementary Figure 1. Heatmap of the correlation between variants in SOX7. rs7005905 and rs7836366, rs10100209 and rs7836366, and rs10100209 and rs7005905 have strong positive linkage disequilibrium (LD) (ρ>0.5); rs4841432 has negative LD with other variants except for rs7009920.

REFERENCES

Abu-Elneel, Kawther, Tsunglin Liu, Francesca S. Gazzaniga, Yuhei Nishimura, Dennis P. Wall, Daniel H. Geschwind, Kaiqin Lao, and Kenneth S. Kosik. 2008. "Heterogeneous dysregulation of microRNAs across the autism spectrum." *neurogenetics* 9 (3): 153-161. https://doi.org/10.1007/s10048-008-0133-5. https://doi.org/10.1007/s10048-008-0133-5.

Alonso-Gonzalez, Aitana, Manuel Calaza, Cristina Rodriguez-Fontenla, and Angel Carracedo. 2019. "Novel Gene-Based Analysis of ASD GWAS: Insight Into the Biological Role of Associated Genes." *Frontiers in Genetics* 10. https://www.frontiersin.org/articles/10.3389/fgene.2019.00733.

Andrews, Simon. 2010. *FastQC: a quality control tool for high throughput sequence data*. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Aspra, Queletzu, Brenda Cabrera-Mendoza, Mirna Edith Morales-Marín, Carla Márquez, Carlos Chicalote, Ana Ballesteros, Miriam Aguilar, Xochitl Castro, Amalia Gómez-Cotero, Ana María Balboa-Verduzco, Lilia Albores-Gallo, Omar Nafate-López, Carlos Alfonso Marcín-Salazar, Patricia Sánchez, Nuria Lanzagorta-Piñol, Fernando Omar López-Armenta, and Humberto Nicolini. 2022. "Epigenome-Wide Analysis Reveals DNA Methylation Alteration in ZFP57 and Its Target RASGFR2 in a Mexican Population Cohort with Autism." *Children* 9 (4): 462. https://doi.org/10.3390/children9040462. https://www.mdpi.com/2227-9067/9/4/462.

Baranova, Ancha, Jun Wang, Hongbao Cao, Jiang-Huan Chen, Jiu Chen, Miao Chen, Sulin Ni, Xijia Xu, Xiaoyan Ke, Shiping Xie, Jing Sun, and Fuquan Zhang. 2022. "Shared genetics between autism spectrum disorder and attention-deficit/hyperactivity disorder and their association with extraversion." *Psychiatry Research* 314: 114679. https://doi.org/10.1016/j.psychres.2022.114679. https://www.sciencedirect.com/science/article/pii/S0165178122002797.

Berto, Stefano, Alex H. Treacher, Emre Caglayan, Danni Luo, Jillian R. Haney, Michael J. Gandal, Daniel H. Geschwind, Albert A. Montillo, and Genevieve Konopka. 2022. "Association between resting-state functional brain connectivity and gene expression is altered in autism spectrum disorder." *Nature Communications* 13 (1): 3328. https://doi.org/10.1038/s41467-022-31053-5. https://www.nature.com/articles/s41467-022-31053-5.

Bjørklund, Geir, Nagwa A. Meguid, Mona A. El-Bana, Alexey A. Tinkov, Khaled Saad, Maryam Dadar, Maha Hemimi, Anatoly V. Skalny, Božena Hosnedlová, Rene Kizek, Joško Osredkar, Mauricio A. Urbina, Teja Fabjan, Amira A. El-Houfey, Joanna Kałużna-Czaplińska, Paulina Gątarek, and Salvatore Chirumbolo. 2020. "Oxidative Stress in Autism Spectrum Disorder." *Molecular Neurobiology* 57 (5): 2314-2332. https://doi.org/10.1007/s12035-019-01742-2. https://doi.org/10.1007/s12035-019-01742-2.

Bourgeron, Thomas. 2015. "From the genetic architecture to synaptic plasticity in autism spectrum disorder." *Nature Reviews Neuroscience* 16 (9): 551-563.

Brandenburg, Cheryl, Jean-Jacques Soghomonian, Kunzhong Zhang, Ina Sulkaj, Brianna Randolph, Marissa Kachadoorian, and Gene J. Blatt. 2020. "Increased Dopamine Type 2 Gene Expression in the Dorsal Striatum in Individuals With Autism Spectrum Disorder Suggests Alterations in Indirect Pathway Signaling and Circuitry." *Frontiers in Cellular Neuroscience* 14. https://www.frontiersin.org/articles/10.3389/fncel.2020.577858.

Brannan, Kris, Hyunmin Kim, Benjamin Erickson, Kira Glover-Cutter, Soojin Kim, Nova Fong, Lauren Kiemele, Kirk Hansen, Richard Davis, Jens Lykke-Andersen, and David L. Bentley. 2012. "mRNA Decapping Factors and the Exonuclease Xrn2 Function in Widespread Premature Termination of RNA Polymerase II Transcription." *Molecular Cell* 46 (3): 311-324. https://doi.org/10.1016/j.molcel.2012.03.006. https://www.sciencedirect.com/science/article/pii/S1097276512002134.

Bryan N. Howie, Peter Donnelly, Jonathan Marchini. 2009. "A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies | PLOS Genetics." https://doi.org/10.1371/journal.pgen.1000529. https://journals.plos.org/plosgenetics/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1000529.

Buitrago, Diana, Mireia Labrador, Juan Pablo Arcon, Rafael Lema, Oscar Flores, Anna Esteve-Codina, Julie Blanc, Nuria Villegas, David Bellido, Marta Gut, Pablo D. Dans, Simon C. Heath, Ivo G. Gut, Isabelle Brun Heath, and Modesto Orozco. 2021. "Impact of DNA methylation on 3D genome structure." *Nature Communications* 12 (1): 3243. https://doi.org/10.1038/s41467-021-23142-8. https://www.nature.com/articles/s41467-021-23142-8.

Buxbaum, Joseph D. 2022. "Multiple rare variants in the etiology of autism spectrum disorders." *Dialogues in clinical neuroscience*.

Caracci, Mario O., Miguel E. Avila, Francisca A. Espinoza-Cavieres, Héctor R. López, Giorgia D. Ugarte, and Giancarlo V. De Ferrari. 2021. "Wnt/β-Catenin-Dependent Transcription in Autism Spectrum Disorders." *Frontiers in Molecular Neuroscience* 14. https://www.frontiersin.org/articles/10.3389/fnmol.2021.764756.

Casanova, Manuel F. 2007. "The Neuropathology of Autism." *Brain Pathology* 17 (4): 422-433. https://doi.org/10.1111/j.1750-3639.2007.00100.x. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1750-3639.2007.00100.x.

Chang, Christopher C, 2071 Stierlin Court Complete Genomics, 94043 Mountain View, CA, USA, Building No. 11 BGI Cognitive Genomics Lab, Bei Shan Industrial Zone, Yantian District, 518083 Shenzhen, China, Carson C Chow, NIDDK/LBM Mathematical Biology Section, National Institutes of Health, 20892 Bethesda, MD, USA, Laurent CAM Tellier, Building No. 11 BGI Cognitive Genomics Lab, Bei Shan Industrial Zone, Yantian District, 518083 Shenzhen, China, University of Copenhagen Bioinformatics Centre, 2200 Copenhagen, Denmark, Shashaank Vattikuti, NIDDK/LBM Mathematical Biology Section, National Institutes of Health, 20892 Bethesda, MD, USA, Shaun M Purcell, Broad Institute of MIT and Harvard Stanley Center for Psychiatric Research, 02142 Cambridge, MA, USA, Department of Psychiatry Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, 10029 New York, NY, USA, Icahn School of Medicine at Mount Sinai Institute for Genomics and Multiscale Biology, 10029 New York, NY, USA, Psychiatric and Neurodevelopmental Genetics Unit Analytic and Translational Genetics Unit, Massachusetts General Hospital, 02114 Boston, MA, USA, James J Lee, NIDDK/LBM Mathematical Biology Section, National Institutes of Health, 20892 Bethesda, MD, USA, and University of Minnesota Twin Cities Department of Psychology, 55455 Minneapolis, MN, USA. 2015. "Second-generation PLINK: rising to the challenge of larger and richer datasets." *GigaScience* 4 (1). https://doi.org/10.1186/s13742-015-0047-8. https://academic.oup.com/gigascience/article-pdf/4/1/s13742-015-0047-8/25512027/13742_2015_article_47.pdf.

Chauhan, Abha, and Ved Chauhan. 2006. "Oxidative stress in autism." *Pathophysiology* 13 (3): 171-181. https://doi.org/10.1016/j.pathophys.2006.05.007. https://www.sciencedirect.com/science/article/pii/S0928468006000538.

Chien, Wei-Hsien, SusanShur-Fen Gau, Chun-Houh Chen, Wen-Che Tsai, Yu-Yu Wu, Po-Hsu Chen, Chi-Yung Shang, and Chia-Hsiang Chen. 2013. "Increased gene expression of FOXP1 in patients with autism spectrum disorders." *Molecular Autism* 4 (1): 23. https://doi.org/10.1186/2040-2392-4-23. https://doi.org/10.1186/2040-2392-4-23.

Consortium, The Autism Spectrum Disorders Working Group of the Psychiatric Genomics. 2017. "Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24. 32 and a significant overlap with schizophrenia." *Molecular autism* 8: 1-17.

Crawley, Jacqueline N., Wolf-Dietrich Heyer, and Janine M. LaSalle. 2016. "Autism and Cancer Share Risk Genes, Pathways, and Drug Targets." *Trends in Genetics* 32 (3): 139-146. https://doi.org/10.1016/j.tig.2016.01.001. https://www.sciencedirect.com/science/article/pii/S0168952516000020.

Crespi, B. 2011. "Autism and cancer risk." *Autism Research* 4 (4): 302-310. https://doi.org/10.1002/aur.208. https://onlinelibrary.wiley.com/doi/abs/10.1002/aur.208.

de la Torre-Ubieta, Luis, Hyejung Won, Jason L. Stein, and Daniel H. Geschwind. 2016. "Advancing the understanding of autism disease mechanisms through genetics." *Nature Medicine* 22 (4): 345-361. https://doi.org/10.1038/nm.4071. https://www.nature.com/articles/nm.4071.

Delaneau, Olivier, Jonathan Marchini, and Jean-François Zagury. 2011. "A linear complexity phasing method for thousands of genomes." *Nature Methods* 9 (2): 179-181. https://doi.org/doi:10.1038/nmeth.1785. https://www.nature.com/articles/nmeth.1785.

Devlin, Bernie, John R. Kelsoe, Pamela Sklar, Mark J. Daly, Michael C. O'Donovan, Nicholas Craddock, Patrick F. Sullivan, Jordan W. Smoller, Kenneth S. Kendler, and others. 2013. "Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs." *Nature genetics* 45 (9): 984-994.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. "STAR: ultrafast universal RNA-seq aligner." *Bioinformatics* 29 (1): 15-21. https://doi.org/10.1093/bioinformatics/bts635. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3530905/.

Eaton, Joshua D., and Steven West. 2018. "An end in sight? Xrn2 and transcriptional termination by RNA polymerase II." *Transcription* 9 (5): 321-326. https://doi.org/10.1080/21541264.2018.1498708. http://www.ncbi.nlm.nih.gov/pubmed/30035655.

El Khouri, E., J. Ghoumid, D. Haye, F. Giuliano, L. Drevillon, A. Briand-Suleau, P. De La Grange, V. Nau, T. Gaillon, T. Bienvenu, H. Jacquemin-Sablon, M. Goossens, S. Amselem, and I. Giurgea. 2021. "Wnt/β-catenin pathway and cell adhesion deregulation in CSDE1-related intellectual disability and autism spectrum disorders." *Molecular Psychiatry* 26 (7): 3572-3585. https://doi.org/10.1038/s41380-021-01072-7. https://www.nature.com/articles/s41380-021-01072-7.

Emanuele, Enzo, Paolo Orsi, Francesco Barale, Stefania Ucelli di Nemi, Marco Bertona, and Pierluigi Politi. 2010. "Serum levels of vascular endothelial growth factor and its receptors in patients with severe autism." *Clinical Biochemistry* 43 (3): 317-319. https://doi.org/10.1016/j.clinbiochem.2009.10.005. https://www.sciencedirect.com/science/article/pii/S0009912009004457.

Ewels, Philip, Måns Magnusson, Sverker Lundin, and Max Käller. 2016. "MultiQC: summarize analysis results for multiple tools and samples in a single report." *Bioinformatics* 32 (19): 3047-3048. https://doi.org/10.1093/bioinformatics/btw354. https://doi.org/10.1093/bioinformatics/btw354.

"FASTX-Toolkit." http://hannonlab.cshl.edu/fastx_toolkit/.

Ferland, Russell J., Timothy J. Cherry, Patricia O. Preware, Edward E. Morrisey, and Christopher A. Walsh. 2003. "Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain." *Journal of Comparative Neurology* 460 (2): 266-279. https://doi.org/10.1002/cne.10654. https://onlinelibrary.wiley.com/doi/abs/10.1002/cne.10654. https://onlinelibrary.wiley.com/doi/full/10.1002/cne.10654.

Fitzgerald, Jonathan B., Moonsoo Jin, Delphine Dean, David J. Wood, Ming H. Zheng, and Alan J. Grodzinsky. 2004. "Mechanical Compression of Cartilage Explants Induces Multiple Time-dependent Gene Expression Patterns and Involves Intracellular Calcium and Cyclic AMP *." *Journal of Biological Chemistry* 279 (19): 19502-19511. https://doi.org/10.1074/jbc.M400437200. https://www.jbc.org/article/S0021-9258(20)67056-9/abstract.

Fombonne, Eric. 2009. "Epidemiology of pervasive developmental disorders." *Pediatric research* 65 (6): 591-598.

Francois, Mathias, Peter Koopman, and Monica Beltrame. 2010. "SoxF genes: Key players in the development of the cardio-vascular system." *The International Journal of Biochemistry & Cell Biology* 42 (3): 445-448. https://doi.org/10.1016/j.biocel.2009.08.017. https://www.sciencedirect.com/science/article/pii/S1357272509002374.

Gabrielli, Alexander P., Ann M. Manzardo, and Merlin G. Butler. 2019. "GeneAnalytics Pathways and Profiling of Shared Autism and Cancer Genes." *International Journal of Molecular Sciences* 20 (5): 1166. https://doi.org/10.3390/ijms20051166. https://www.mdpi.com/1422-0067/20/5/1166.

Gaugler, Trent, Lambertus Klei, Stephan J. Sanders, Corneliu A. Bodea, Arthur P. Goldberg, Ann B. Lee, Milind Mahajan, Dina Manaa, Yudi Pawitan, Jennifer Reichert, and others. 2014. "Most genetic risk for autism resides with common variation." *Nature genetics* 46 (8): 881-885.

Ghahramani Seno, Mohammad M., Pingzhao Hu, Fuad G. Gwadry, Dalila Pinto, Christian R. Marshall, Guillermo Casallo, and Stephen W. Scherer. 2011. "Gene and miRNA expression profiles in autism spectrum disorders." *Brain Research* 1380: 85-97. https://doi.org/10.1016/j.brainres.2010.09.046. https://www.sciencedirect.com/science/article/pii/S0006899310020512. https://www.sciencedirect.com/science/article/abs/pii/S0006899310020512.

Gillberg, Christopher, Elisabeth Fernell, and Helen Minnis. 2014. Early symptomatic syndromes eliciting neurodevelopmental clinical examinations. Hindawi.

Gozal, Evelyne, Rekha Jagadapillai, Jun Cai, and Gregory N. Barnes. 2021. "Potential crosstalk between sonic hedgehog-WNT signaling and neurovascular molecules: Implications for blood–brain barrier integrity in autism spectrum disorder." *Journal of Neurochemistry* 159 (1): 15-28. https://doi.org/10.1111/jnc.15460. https://onlinelibrary.wiley.com/doi/abs/10.1111/jnc.15460.

Grove, Jakob, Stephan Ripke, Thomas D. Als, Manuel Mattheisen, Raymond K. Walters, Hyejung Won, Jonatan Pallesen, Esben Agerbo, Ole A. Andreassen, Richard Anney, and others. 2019. "Identification of common genetic risk variants for autism spectrum disorder." *Nature genetics* 51 (3): 431-444.

Guo, Bin, and Baolin Wu. 2018. "Statistical methods to detect novel genetic variants using publicly available GWAS summary data." *Computational Biology and Chemistry* 74: 76-79. https://doi.org/https://doi.org/10.1016/j.compbiolchem.2018.02.016. https://www.sciencedirect.com/science/article/pii/S1476927117306953.

Hannon, Eilis, Diana Schendel, Christine Ladd-Acosta, Jakob Grove, Esben Agerbo, Thomas D. Als, Rich Belliveau, Jonas Bybjerg-Grauholm, Marie Bækved-Hansen, Anders Børglum, Felecia Cerrato, Jane Christensen, Kimberly Chambert, Claire Churchhouse, Mark Daly, Ditte Demontis, Ashley Dumont, Jacqueline Goldstein, Christine Hansen, Mads Hauberg, David Hougaard, Daniel Howrigan, Hailiang Huang, Julian Maller, Alicia Martin, Joanna Martin, Manuel Mattheisen, Jennifer Moran, Ole Mors, Preben Mortensen, Benjamin Neale, Merete Nordentoft, Mette Nyegaard, Jonatan Pallsen, Duncan Palmer, Carsten Pedersen, Marianne Pedersen, Timothy Poterba, Jesper Poulsen, Per Qvist, Stephan Ripke, Elise Robinson, Kyle Satterstrom, Christine Stevens, Patrick Turley, Raymond Walters, Thomas Werge, Christine Søholm Hansen, Shan V. Andrews, David Michael Hougaard, Michaeline Bresnahan, Mads Vilhelm Hollegaard, Marie Bækvad-Hansen, Mady Hornig, Preben Bo Mortensen, Anders D. Børglum, Marianne Giørtz Pedersen, Joseph Buxbaum, M. Daniele Fallin, Abraham Reichenberg, Jonathan Mill, and Psych-Broad A. S. D. Group i. 2018. "Elevated polygenic burden for autism is associated with differential DNA methylation at birth." *Genome Medicine* 10 (1): 19. https://doi.org/10.1186/s13073-018-0527-4. https://doi.org/10.1186/s13073-018-0527-4. https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-018-0527-4.

Hicks, Steven D., and Frank A. Middleton. 2016. "A Comparative Review of microRNA Expression Patterns in Autism Spectrum Disorder." *Frontiers in Psychiatry* 7. https://www.frontiersin.org/articles/10.3389/fpsyt.2016.00176.

Hormozdiari, Fereydoun, Osnat Penn, Elhanan Borenstein, and Evan E. Eichler. 2015. "The discovery of integrated gene networks for autism and related disorders." *Genome Research* 25 (1): 142-154. https://doi.org/10.1101/gr.178855.114. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4317170/.

Hsieh, Adam H., Cliff M. H. Tsai, Qing-Jun Ma, Tong Lin, Albert J. Banes, Francisco J. Villarreal, Wayne H. Akeson, and K. L. Paul Sung. 2000. "Time-dependent increases in type-III collagen gene expression in medial collateral ligament fibroblasts under cyclic strains." *Journal of Orthopaedic Research* 18 (2): 220-227. https://doi.org/10.1002/jor.1100180209. https://onlinelibrary.wiley.com/doi/abs/10.1002/jor.1100180209.

Huang, Kunling, Yuchang Wu, Junha Shin, Ye Zheng, Alireza Fotuhi Siahpirani, Yupei Lin, Zheng Ni, Jiawen Chen, Jing You, Sunduz Keles, Daifeng Wang, Sushmita Roy, and Qiongshi Lu. 2021. "Transcriptome-wide transmission disequilibrium analysis identifies novel risk genes for autism spectrum disorder." *PLOS Genetics* 17 (2): e1009309. https://doi.org/10.1371/journal.pgen.1009309. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1009309.

I, Voineagu, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, and Geschwind DH. 2011. "Transcriptomic analysis of autistic brain reveals convergent molecular pathology." *Nature* 474 (7351). https://doi.org/10.1038/nature10110. https://www.ncbi.nlm.nih.gov/pubmed/21614001.

Jamain, Stéphane, Catalina Betancur, Hélène Quach, Anne Philippe, Marc Fellous, Bruno Giros, Christopher Gillberg, Marion Leboyer, and Thomas Bourgeron. 2002. "Linkage and association of the glutamate receptor 6 gene with autism." *Molecular psychiatry* 7 (3): 302-310.

Jamain, Stéphane, Hélène Quach, Catalina Betancur, Maria Råstam, Catherine Colineaux, I. Carina Gillberg, Henrik Soderstrom, Bruno Giros, Marion Leboyer, Christopher Gillberg, and others. 2003. "Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism." *Nature genetics* 34 (1): 27-29.

Johnson, F. Brad. 2011. "PinX1 the tail on the chromosome." *The Journal of Clinical Investigation* 121 (4): 1242-1244. https://doi.org/10.1172/JCI57024. https://www.jci.org/articles/view/57024. http://www.ncbi.nlm.nih.gov/pubmed/21436580.

Katoh, Masaru. 2002. "Expression of human SOX7 in normal tissues and tumors." *International Journal of Molecular Medicine* 9 (4): 363-368. https://doi.org/10.3892/ijmm.9.4.363. https://www.spandidos-publications.com/10.3892/ijmm.9.4.363.

Keshet, Ilana, Judy Lieman-Hurwitz, and Howard Cedar. 1986. "DNA methylation affects the formation of active chromatin." *Cell* 44 (4): 535-543. https://doi.org/10.1016/0092-8674(86)90263-1. https://www.sciencedirect.com/science/article/pii/0092867486902631. https://www.sciencedirect.com/science/article/abs/pii/0092867486902631.

Kim, Il-Kug, Kangsan Kim, Eunhyeong Lee, Dong Sun Oh, Chan Soon Park, Seongyeol Park, Jee Myung Yang, Ju-Hee Kim, Hyung-Seok Kim, David T. Shima, Jeong Hoon Kim, Seok Ho Hong, Young Hyun Cho, Young Hoon Kim, Jong Bae Park, Gou Young Koh, Young Seok Ju, Heung Kyu Lee, Seungjoo Lee, and Injune Kim. 2018. "Sox7 promotes high-grade glioma by increasing VEGFR2-mediated vascular abnormality." *The Journal of Experimental Medicine* 215 (3): 963-983. https://doi.org/10.1084/jem.20170123. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5839752/.

Kim, Kangsan, Il-Kug Kim, Jee Myung Yang, Eunhyeong Lee, Bong Ihn Koh, Sukhyun Song, Junseong Park, Sungsu Lee, Chulhee Choi, Jin Woo Kim, Yoshiaki Kubota, Gou Young Koh, and Injune Kim. 2016. "SoxF Transcription Factors Are Positive Feedback Regulators of VEGF Signaling." *Circulation Research* 119 (7): 839-852. https://doi.org/10.1161/CIRCRESAHA.116.308483. https://www.ahajournals.org/doi/full/10.1161/CIRCRESAHA.116.308483.

Kinjo, Erika Reime, Guilherme Shigueto Vilar Higa, Erica de Sousa, Otávio Augusto Nocera Casado, Marcio Vinicius Damico, Luiz Roberto G. Britto, and Alexandre Hiroaki Kihara. 2013. "A possible new mechanism for the control of miRNA expression in neurons." *Experimental Neurology* 248: 546-558. https://doi.org/10.1016/j.expneurol.2013.07.022. https://www.sciencedirect.com/science/article/pii/S0014488613002392.

Klomp, Jeff, James Hyun, Jennifer E. Klomp, Kostandin Pajcini, Jalees Rehman, and Asrar B. Malik. 2020. "Comprehensive transcriptomic profiling reveals SOX7 as an early regulator of angiogenesis in hypoxic human endothelial cells." *Journal of Biological Chemistry* 295 (15): 4796-4808. https://doi.org/10.1074/jbc.RA119.011822. https://www.jbc.org/article/S0021-9258(17)48577-2/abstract. http://www.ncbi.nlm.nih.gov/pubmed/32071080.

Kwan, Vickie, Brianna K. Unda, and Karun K. Singh. 2016. "Wnt signaling networks in autism spectrum disorder and intellectual disability." *Journal of Neurodevelopmental Disorders* 8 (1): 1-10.

https://doi.org/10.1186/s11689-016-9176-3. https://jneurodevdisorders.biomedcentral.com/articles/10.1186/s11689-016-9176-3.

Lawlor, Nathan, Joshy George, Mohan Bolisetty, Romy Kursawe, Lili Sun, V. Sivakamasundari, Ina Kycia, Paul Robson, and Michael L. Stitzel. 2017. "Single-cell transcriptomes identify human islet cell signatures and reveal cell-type–specific expression changes in type 2 diabetes." *Genome Research* 27 (2): 208-222. https://doi.org/10.1101/gr.212720.116. https://genome.cshlp.org/content/27/2/208. http://www.ncbi.nlm.nih.gov/pubmed/27864352.

Leblond, Claire S., Freddy Cliquet, Coralie Carton, Guillaume Huguet, Alexandre Mathieu, Thomas Kergrohen, Julien Buratti, Nathalie Lemière, Laurence Cuisset, Thierry Bienvenu, and others. 2019. "Both rare and common genetic variants contribute to autism in the Faroe Islands." *NPJ genomic medicine* 4 (1): 1-10.

Levy, Dan, Michael Ronemus, Boris Yamrom, Yoon-ha Lee, Anthony Leotta, Jude Kendall, Steven Marks, B. Lakshmi, Deepa Pai, Kenny Ye, and others. 2011. "Rare de novo and transmitted copy-number variation in autistic spectrum disorders." *Neuron* 70 (5): 886-897.

Lewis, Joe, and Adrian Bird. 1991. "DNA methylation and chromatin structure." *FEBS Letters* 285 (2): 155-159. https://doi.org/10.1016/0014-5793(91)80795-5. https://onlinelibrary.wiley.com/doi/abs/10.1016/0014-5793%2891%2980795-5. https://febs.onlinelibrary.wiley.com/doi/abs/10.1016/0014-5793(91)80795-5.

Li, Xiaohong, Hua Zou, and W. Ted Brown. 2012. "Genes associated with autism spectrum disorder." *Brain research bulletin* 88 (6): 543-552.

Lilly, Andrew J., Andrzej Mazan, Daryl A. Scott, Georges Lacaud, and Valerie Kouskoff. 2017. "SOX7 expression is critically required in FLK1-expressing cells for vasculogenesis and angiogenesis during mouse embryonic development." *Mechanisms of Development* 146: 31-41. https://doi.org/10.1016/j.mod.2017.05.004. https://www.sciencedirect.com/science/article/pii/S0925477317300096.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology* 15 (12): 550. https://doi.org/10.1186/s13059-014-0550-8. https://doi.org/10.1186/s13059-014-0550-8. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8.

MacDonald, Bryan T., Keiko Tamai, and Xi He. 2009. "Wnt/β-Catenin Signaling: Components, Mechanisms, and Diseases." *Developmental Cell* 17 (1): 9-26. https://doi.org/10.1016/j.devcel.2009.06.016. https://www.sciencedirect.com/science/article/pii/S1534580709002573.

Marques, Ana Rita, João Xavier Santos, Hugo Martiniano, Joana Vilela, Célia Rasga, Luísa Romão, and Astrid Moura Vicente. 2022. "Gene Variants Involved in Nonsense-Mediated mRNA Decay Suggest a Role in Autism Spectrum Disorder." *Biomedicines* 10 (3): 665. https://doi.org/10.3390/biomedicines10030665. https://www.mdpi.com/2227-9059/10/3/665.

Melke, Jonas, Hany Goubran Botros, Pauline Chaste, Catalina Betancur, Gudrun Nygren, Henrik Anckarsäter, Maria Rastam, Ola Ståhlberg, I. Carina Gillberg, Richard Delorme, and others. 2008. "Abnormal melatonin synthesis in autism spectrum disorders." *Molecular psychiatry* 13 (1): 90-98.

Nagarajan, Vinay K., Christopher I. Jones, Sarah F. Newbury, and Pamela J. Green. 2013. "XRN 5′→3′ exoribonucleases: Structure, mechanisms and functions." *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1829 (6): 590-603. https://doi.org/10.1016/j.bbagrm.2013.03.005. https://www.sciencedirect.com/science/article/pii/S1874939913000503. https://www.sciencedirect.com/science/article/abs/pii/S1874939913000503.

Nussbacher, Julia K., Ricardos Tabet, Gene W. Yeo, and Clotilde Lagier-Tourenne. 2019. "Disruption of RNA Metabolism in Neurological Diseases and Emerging Therapeutic Interventions." *Neuron* 102 (2): 294-320. https://doi.org/10.1016/j.neuron.2019.03.014. https://www.sciencedirect.com/science/article/pii/S0896627319302776.

Oshimori, Naoki, Miho Ohsugi, and Tadashi Yamamoto. 2006. "The Plk1 target Kizuna stabilizes mitotic centrosomes to ensure spindle bipolarity." *Nature Cell Biology* 8 (10): 1095-1101. https://doi.org/10.1038/ncb1474. http://www.ncbi.nlm.nih.gov/pubmed/16980960.

Ouellette, Julie, Xavier Toussay, Cesar H. Comin, Luciano da F. Costa, Mirabelle Ho, María Lacalle-Aurioles, Moises Freitas-Andrade, Qing Yan Liu, Sonia Leclerc, Youlian Pan, Ziying Liu, Jean-François Thibodeau, Melissa Yin, Micael Carrier, Cameron J. Morse, Peter Van Dyken, Christopher J. Bergin, Sylvain Baillet, Christopher R. Kennedy, Marie-Ève Tremblay, Yannick D. Benoit, William L. Stanford, Dylan Burger, Duncan J. Stewart, and Baptiste Lacoste. 2020. "Vascular contributions to 16p11.2 deletion autism syndrome modeled in mice." *Nature Neuroscience* 23 (9): 1090-1101. https://doi.org/10.1038/s41593-020-0663-1. https://www.nature.com/articles/s41593-020-0663-1.

O'Roak, Brian J., Laura Vives, Santhosh Girirajan, Emre Karakoc, Niklas Krumm, Bradley P. Coe, Roie Levy, Arthur Ko, Choli Lee, Joshua D. Smith, and others. 2012. "Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations." *Nature* 485 (7397): 246-250.

Packer, Alan. 2016. "Neocortical neurogenesis and the etiology of autism spectrum disorder." *Neuroscience & Biobehavioral Reviews* 64: 185-195. https://doi.org/10.1016/j.neubiorev.2016.03.002. https://www.sciencedirect.com/science/article/pii/S0149763416300355. https://www.sciencedirect.com/science/article/pii/S0149763416300355#sec0010.

Pain, Oliver, Andrew J. Pocklington, Peter A. Holmans, Nicholas J. Bray, Heath E. O'Brien, Lynsey S. Hall, Antonio F. Pardiñas, Michael C. O'Donovan, Michael J. Owen, and Richard Anney. 2019. "Novel Insight Into the Etiology of Autism Spectrum Disorder Gained by Integrating Expression Data With Genome-wide Association Statistics." *Biological Psychiatry* 86 (4): 265-273. https://doi.org/10.1016/j.biopsych.2019.04.034. https://www.sciencedirect.com/science/article/pii/S0006322319313344.

Parikshak, Neelroop N., Vivek Swarup, T. Grant Belgard, Manuel Irimia, Gokul Ramaswami, Michael J. Gandal, Christopher Hartl, Virpi Leppa, Luis de la Torre Ubieta, Jerry Huang, Jennifer K. Lowe, Benjamin J. Blencowe, Steve Horvath, and Daniel H. Geschwind. 2016. "Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism." *Nature* 540 (7633): 423-427. https://doi.org/10.1038/nature20612. https://www.nature.com/articles/nature20612.

Peyre, Hugo, Tabea Schoeler, Chaoyu Liu, Camille Michèle Williams, Nicolas Hoertel, Alexandra Havdahl, and Jean-Baptiste Pingault. 2020. Combining multivariate genomic approaches to elucidate the comorbidity between ASD and ADHD. bioRxiv.

Pramparo, Tiziano, Michael V. Lombardo, Kathleen Campbell, Cynthia Carter Barnes, Steven Marinero, Stephanie Solso, Julia Young, Maisi Mayo, Anders Dale, Clelia Ahrens-Barbeau, Sarah S. Murray, Linda Lopez, Nathan Lewis, Karen Pierce, and Eric Courchesne. 2015. "Cell cycle networks link gene expression dysregulation, mutation, and brain maldevelopment in autistic toddlers." *Molecular Systems Biology* 11 (12): 841. https://doi.org/10.15252/msb.20156108. https://www.embopress.org/doi/full/10.15252/msb.20156108.

Putri, Givanna H., Simon Anders, Paul Theodor Pyl, John E. Pimanda, and Fabio Zanini. 2022. "Analysing high-throughput sequencing data in Python with HTSeq 2.0." *Bioinformatics* 38 (10): 2943-2945. https://doi.org/10.1093/bioinformatics/btac166. https://doi.org/10.1093/bioinformatics/btac166. https://academic.oup.com/bioinformatics/article/38/10/2943/6551247?login=false.

Páez, Marco T., Toshiyuki Yamamoto, Ken-ichi Hayashi, Toshiyuki Yasuda, Naoki Harada, Naomichi Matsumoto, Kenji Kurosawa, Yoshiyuki Furutani, Shuichi Asakawa, Nobuyoshi Shimizu, and Rumiko Matsuoka. 2008. "Two patients with atypical interstitial deletions of 8p23.1: Mapping of phenotypical traits." *American Journal of Medical Genetics Part A* 146A (9): 1158-1165. https://doi.org/10.1002/ajmg.a.32205. https://onlinelibrary.wiley.com/doi/abs/10.1002/ajmg.a.32205.

Quesnel-Vallières, Mathieu, Robert J. Weatheritt, Sabine P. Cordes, and Benjamin J. Blencowe. 2019. "Autism spectrum disorder: insights into convergent mechanisms from transcriptomics." *Nature*

Reviews Genetics 20 (1): 51-63. https://doi.org/10.1038/s41576-018-0066-2. https://www.nature.com/articles/s41576-018-0066-2.

Rahman, Md Rezanur, Maria Cristina Petralia, Rosella Ciurleo, Alessia Bramanti, Paolo Fagone, Md Shahjaman, Lang Wu, Yanfa Sun, Beste Turanli, Kazim Yalcin Arga, Md Rafiqul Islam, Tania Islam, and Ferdinando Nicoletti. 2020. "Comprehensive Analysis of RNA-Seq Gene Expression Profiling of Brain Transcriptomes Reveals Novel Genes, Regulators, and Pathways in Autism Spectrum Disorder." *Brain Sciences* 10 (10): 747. https://doi.org/10.3390/brainsci10100747. https://www.mdpi.com/2076-3425/10/10/747.

Reis, Marco, and Stefan Liebner. 2013. "Wnt signaling in the vasculature." *Experimental Cell Research* 319 (9): 1317-1323. https://doi.org/10.1016/j.yexcr.2012.12.023. https://www.sciencedirect.com/science/article/pii/S0014482712005009.

Robinson, Elise B., Beate St Pourcain, Verneri Anttila, Jack A. Kosmicki, Brendan Bulik-Sullivan, Jakob Grove, Julian Maller, Kaitlin E. Samocha, Stephan J. Sanders, Stephan Ripke, and others. 2016. "Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population." *Nature genetics* 48 (5): 552-555.

Roohi, Jasmin, Cristina Montagna, David H. Tegay, Lance E. Palmer, Carla DeVincent, John C. Pomeroy, Susan L. Christian, Norma Nowak, and Eli Hatchwell. 2009. "Disruption of contactin 4 in three subjects with autism spectrum disorder." *Journal of medical genetics* 46 (3): 176-182.

Sanders, Stephan J., A. Gulhan Ercan-Sencicek, Vanessa Hus, Rui Luo, Michael T. Murtha, Daniel Moreno-De-Luca, Su H. Chu, Michael P. Moreau, Abha R. Gupta, Susanne A. Thomson, and others. 2011. "Multiple recurrent de novo CNVs, including duplications of the 7q11. 23 Williams syndrome region, are strongly associated with autism." *Neuron* 70 (5): 863-885.

Sanders, Stephan J., Xin He, A. Jeremy Willsey, A. Gulhan Ercan-Sencicek, Kaitlin E. Samocha, A. Ercument Cicek, Michael T. Murtha, Vanessa H. Bal, Somer L. Bishop, Shan Dong, and others. 2015. "Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci." *Neuron* 87 (6): 1215-1233.

Sanders, Stephan J., Michael T. Murtha, Abha R. Gupta, John D. Murdoch, Melanie J. Raubeson, A. Jeremy Willsey, A. Gulhan Ercan-Sencicek, Nicholas M. DiLullo, Neelroop N. Parikshak, Jason L. Stein, and others. 2012. "De novo mutations revealed by whole-exome sequencing are strongly associated with autism." *Nature* 485 (7397): 237-241.

Sandin, Sven, Paul Lichtenstein, Ralf Kuja-Halkola, Christina Hultman, Henrik Larsson, and Abraham Reichenberg. 2017. "The heritability of autism spectrum disorder." *Jama* 318 (12): 1182-1184.

Satterstrom, F. Kyle, Jack A. Kosmicki, Jiebiao Wang, Michael S. Breen, Silvia De Rubeis, Joon-Yong An, Minshi Peng, Ryan Collins, Jakob Grove, Lambertus Klei, and others. 2020. "Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism." *Cell* 180 (3): 568-584.

Shen-Orr, Shai S., Robert Tibshirani, Purvesh Khatri, Dale L. Bodian, Frank Staedtler, Nicholas M. Perry, Trevor Hastie, Minnie M. Sarwal, Mark M. Davis, and Atul J. Butte. 2010. "Cell type–specific gene expression differences in complex tissues." *Nature Methods* 7 (4): 287-289. https://doi.org/10.1038/nmeth.1439. https://www.nature.com/articles/nmeth.1439.

"SRA Toolkit." *GitHub*. https://github.com/ncbi/sra-tools/wiki/01.-Downloading-SRA-Toolkit. https://github.com/ncbi/sra-tools.

Stenman, Jan M., Jay Rajagopal, Thomas J. Carroll, Makoto Ishibashi, Jill McMahon, and Andrew P. McMahon. 2008. "Canonical Wnt Signaling Regulates Organ-Specific Assembly and Differentiation of CNS Vasculature." *Science* 322 (5905): 1247-1250. https://doi.org/10.1126/science.1164594. https://www.science.org/doi/full/10.1126/science.1164594.

Tabarés-Seisdedos, R., and J. L. R. Rubenstein. 2009. "Chromosome 8p as a potential hub for developmental neuropsychiatric disorders: implications for schizophrenia, autism and cancer." *Molecular Psychiatry* 14 (6): 563-589. https://doi.org/10.1038/mp.2009.2. https://www.nature.com/articles/mp20092.

Takash, W., J. Cañizares, N. Bonneaud, F. Poulat, M. G. Mattéi, P. Jay, and P. Berta. 2001. "SOX7 transcription factor: sequence, chromosomal localisation, expression, transactivation and interference with Wnt signalling." *Nucleic Acids Research* 29 (21): 4274-4283. https://doi.org/10.1093/nar/29.21.4274. http://www.ncbi.nlm.nih.gov/pubmed/11691915.

Tick, Beata, Patrick Bolton, Francesca Happé, Michael Rutter, and Frühling Rijsdijk. 2016. "Heritability of autism spectrum disorders: a meta-analysis of twin studies." *Journal of Child Psychology and Psychiatry* 57 (5): 585-595.

Vallée, Alexandre, Jean-Noël Vallée, and Yves Lecarpentier. 2019. "PPARγ agonists: potential treatment for autism spectrum disorder by inhibiting the canonical WNT/β-catenin pathway." *Molecular Psychiatry* 24 (5): 643-652. https://doi.org/10.1038/s41380-018-0131-4. https://www.nature.com/articles/s41380-018-0131-4.

Wang, Minshi, and Dimitri G. Pestov. 2011. "5′-end surveillance by Xrn2 acts as a shared mechanism for mammalian pre-rRNA maturation and decay." *Nucleic Acids Research* 39 (5): 1811-1822. https://doi.org/10.1093/nar/gkq1050. https://doi.org/10.1093/nar/gkq1050.

Wat, Margaret J., Oleg A. Shchelochkov, Ashley M. Holder, Amy M. Breman, Aditi Dagli, Carlos Bacino, Fernando Scaglia, Roberto T. Zori, Sau Wai Cheung, Daryl A. Scott, and Sung-Hae Lee Kang. 2009. "Chromosome 8p23.1 deletions as a cause of complex congenital heart defects and diaphragmatic hernia." *American Journal of Medical Genetics Part A* 149A (8): 1661-1677. https://doi.org/10.1002/ajmg.a.32896. https://onlinelibrary.wiley.com/doi/abs/10.1002/ajmg.a.32896.

Weber, Axel, Angelika Köhler, Andreas Hahn, and Ulrich Müller. 2014. "8p23.1 duplication syndrome: narrowing of critical interval to 1.80 Mbp." *Molecular Cytogenetics* 7 (1): 94. https://doi.org/10.1186/s13039-014-0094-3. https://doi.org/10.1186/s13039-014-0094-3.

West, Steven, Natalia Gromak, and Nick J. Proudfoot. 2004. "Human 5′ → 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites." *Nature* 432 (7016): 522-525. https://doi.org/10.1038/nature03035. https://www.nature.com/articles/nature03035.

Weyer, Anja, and Karl Schilling. 2003. "Developmental and cell type-specific expression of the neuronal marker NeuN in the murine cerebellum." *Journal of Neuroscience Research* 73 (3): 400-409. https://doi.org/10.1002/jnr.10655. https://onlinelibrary.wiley.com/doi/abs/10.1002/jnr.10655.

Wu, Ye E., Neelroop N. Parikshak, T. Grant Belgard, and Daniel H. Geschwind. 2016. "Genome-wide, integrative analysis implicates microRNA dysregulation in autism spectrum disorder." *Nature Neuroscience* 19 (11): 1463-1476. https://doi.org/10.1038/nn.4373. https://www.nature.com/articles/nn.4373.

Xu, Xiaoxiao, Alan B. Wells, David R. O'Brien, Arye Nehorai, and Joseph D. Dougherty. 2014. "Cell Type-Specific Expression Analysis to Identify Putative Cellular Mechanisms for Neurogenetic Disorders." *Journal of Neuroscience* 34 (4): 1420-1431. https://doi.org/10.1523/JNEUROSCI.4488-13.2014. https://www.jneurosci.org/content/34/4/1420. http://www.ncbi.nlm.nih.gov/pubmed/24453331.

Yang, Zhiyu, Hanrui Wu, Phil H. Lee, Fotis Tsetsos, Lea K. Davis, Dongmei Yu, Sang Hong Lee, Søren Dalsgaard, Jan Haavik, Csaba Barta, Tetyana Zayats, Valsamma Eapen, Naomi R. Wray, Bernie Devlin, Mark Daly, Benjamin Neale, Anders D. Børglum, James J. Crowley, Jeremiah Scharf, Carol A. Mathews, Stephen V. Faraone, Barbara Franke, Manuel Mattheisen, Jordan W. Smoller, and Peristera Paschou. 2021. "Investigating Shared Genetic Basis Across Tourette Syndrome and Comorbid Neurodevelopmental Disorders Along the Impulsivity-Compulsivity Spectrum." *Biological Psychiatry* 90 (5): 317-327. https://doi.org/10.1016/j.biopsych.2020.12.028. https://www.sciencedirect.com/science/article/pii/S000632232100038X.

Yao, Yuemang, William J. Walsh, Woody R. McGinnis, and Domenico Praticò. 2006. "Altered Vascular Phenotype in Autism: Correlation With Oxidative Stress." *Archives of Neurology* 63 (8): 1161-1164. https://doi.org/10.1001/archneur.63.8.1161. https://doi.org/10.1001/archneur.63.8.1161.

Yasuda, Yuka, Ryota Hashimoto, Hidenaga Yamamori, Kazutaka Ohi, Motoyuki Fukumoto, Satomi Umeda-Yano, Ikuko Mohri, Akira Ito, Masako Taniike, and Masatoshi Takeda. 2011. "Gene expression analysis in lymphoblasts derived from patients with autism spectrum disorder." *Molecular Autism* 2 (1): 9. https://doi.org/10.1186/2040-2392-2-9. https://doi.org/10.1186/2040-2392-2-9.

Zhang, Yinghua, Yan Sun, Fei Wang, Zhongping Wang, Yuwen Peng, and Ruixi Li. 2012. "Downregulating the Canonical Wnt/β-catenin Signaling Pathway Attenuates the Susceptibility to Autism-like Phenotypes by Decreasing Oxidative Stress." *Neurochemical Research* 37 (7): 1409-1419. https://doi.org/10.1007/s11064-012-0724-2. https://doi.org/10.1007/s11064-012-0724-2.

Zhao, Tianshu, Hui Yang, Yu Tian, Qing Xie, Yun Lu, Yu Wang, Ning Su, Baijing Dong, Xian Liu, Ce Wang, Chuanlu Jiang, and Xiaoqian Liu. 2016. "SOX7 is associated with the suppression of human glioma by HMG-box dependent regulation of Wnt/β-catenin signaling." *Cancer Letters* 375 (1): 100-107. https://doi.org/10.1016/j.canlet.2016.02.044. http://www.ncbi.nlm.nih.gov/pubmed/26944317.

FUNDING SUPPORT

This study was financially supported by the National Institutes of Health (NCATS R44TR003491 and NIDDK UH3DK119982) and the University of North Texas (Startup).